

Faculty of Health Sciences

Basal statistik

Logaritmer, Repetition, Kovariansanalyse, Interaktion, i SAS

Lene Theil Skovgaard

28. september 2020



Logaritmer og kovariansanalyse

- ▶ Parret sammenligning af målemetoder, med logaritmer
- ▶ Kovariansanalyse: Confounding og mediering
- ▶ Sammenligning af hældninger
Interaktion - igen
- ▶ Kovariansanalyse:
Eksempel om baseline og follow-up

E-mail: 1tsk@sund.ku.dk

*: Siden er lidt teknisk



Sammenligning af målemetoder

(jvf. eksemplet MF vs. SV fra første forelæsning)

To forskellige metoder til bestemmelse af glucosekoncentration.

Ref: R.G. Miller et.al. (eds): Biostatistics Casebook. Wiley, 1980

REFE:

Farvetest, der kan 'forurenes' af urinsyre

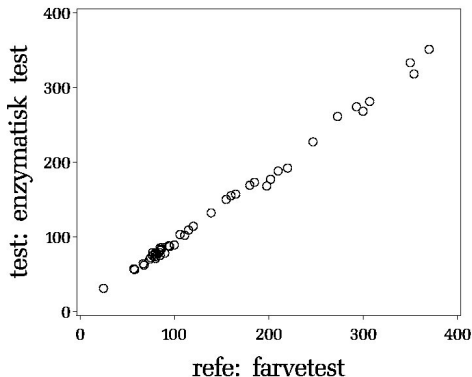
TEST:

Enzymatisk test, mere specifikt for glucose.

<i>nr.</i>	<i>REFE</i>	<i>TEST</i>
1	155	150
2	160	155
3	180	169
.	.	.
.	.	.
.	.	.
44	94	88
45	111	102
46	210	188
$\bar{X}$	144.1	134.2
SD	91.0	83.2



Scatter plot af de to metoder



Det ser jo pænt ud...., eller?

Formål med undersøgelsen

kunne være:

- ▶ Kan vi erstatte en dårlig metode med en bedre metode?
- ▶ Kan vi erstatte en dyr metode med en billigere metode?
- ▶ Kan vi bruge de to metoder i flæng?

Under alle omstændigheder:

Hvor store forskelle kan vi forvente at finde mellem de to metoder

Og svaret er **ikke**:

Korrelationen er 0.99776.....



Parret sammenligning

Det er naturligt at se på differenser (dif=refe-test)

- ▶ Test om middelværdien kan være 0:
Systematisk forskel? (denne side)
- ▶ Kvantificer individuelle differenser:
Limits of agreement (næste side)

The MEANS Procedure

Variable	N	Mean	Std Dev	t Value	Pr > t
dif	46	9.8913043	9.7027562	6.91	<.0001

Med en P-værdi < 0.0001 er der **stærk indikation** af forskel på de to metoder



Limits of agreement

På basis af en normalfordelingsantagelse på differenserne finder vi referenceintervallet (normalområdet):

$$9.89 \pm 2 \times 9.70 = (-9.51, 29.29)$$

med fortolkningen:

“Når vi måler med begge metoder på samme person, vil differensen typisk ligge i intervallet $(-9.5, 29.3)$ ”

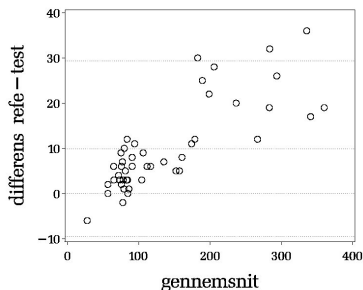
På tegningen ses, at dette er en **dårlig** beskrivelse, idet

- ▶ differenserne stiger med niveauet (repræsenteret ved gennemsnittet)
- ▶ variationen stiger også med niveauet



Bland-Altman plot

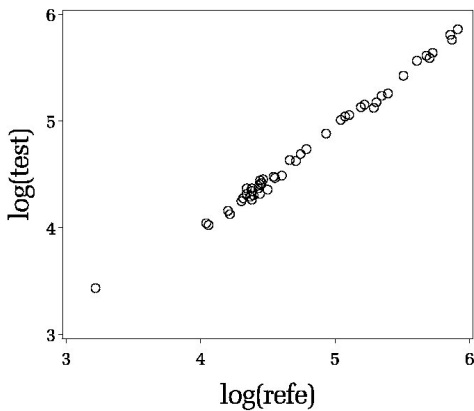
Plot af differenser mod gennemsnit (af de to målinger på samme person), se kode s. 82



Store afvigelser ved høje målinger, dvs. **relative afvigelser**
Så er det smart at se på **logaritmer**

Scatter plot

efter logaritmetransformation (her “den naturlige”)

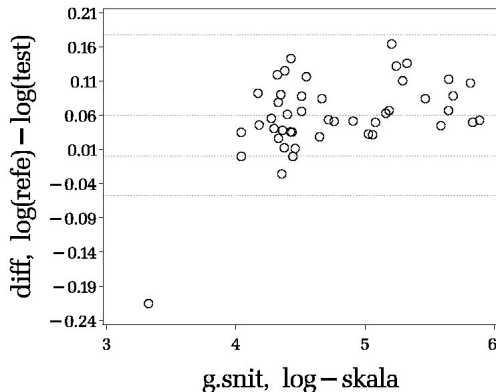


Bemærk

- ▶ Det er de **oprindelige målinger**, der skal logaritmetransformeres, *ikke* differenserne!
- ▶ Det er ligegyldigt, hvilken logaritmefunktion, der vælges (der er proportionalitet mellem alle logaritmer)
- ▶ Efter logaritmering gentages proceduren med differenser og konstruktion af limits of agreement



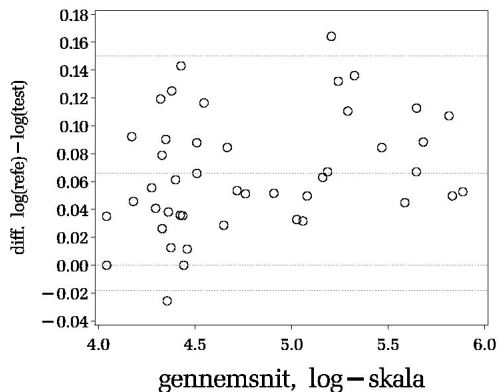
Bland-Altman plot for logaritmer



Der er en tydelig *outlier* (den mindste observation)

Vi udelader en outlier....

og laver igen et Bland-Altman plot



som bliver acceptabelt....

En slags konklusion

The MEANS Procedure

Variable	N	Mean	Std Dev	t Value	Pr > t
ldif	45	0.0657295	0.0419547	10.51	<.0001

Limits of agreement på logaritmisk skala:

$$0.066 \pm 2 \times 0.042 = (-0.018, 0.150)$$

Det betyder, at der i 95% af tilfældene vil gælde

$$-0.018 < \log(\text{REFE}) - \log(\text{TEST}) = \log\left(\frac{\text{REFE}}{\text{TEST}}\right) < 0.150$$

Men hvad kan vi bruge det til?



Konklusion på brugbar facon

Vi kan **tilbagetransformere** med **anti-logaritmen** (her er det eksponentialfunktionen $\exp()$) og få

$$\exp(-0.018) = 0.982 < \frac{\text{REFE}}{\text{TEST}} < 1.162 = \exp(0.150) \quad \text{eller 'omvendt'}$$

$$0.861 < \frac{\text{TEST}}{\text{REFE}} < 1.018$$

Det betyder:

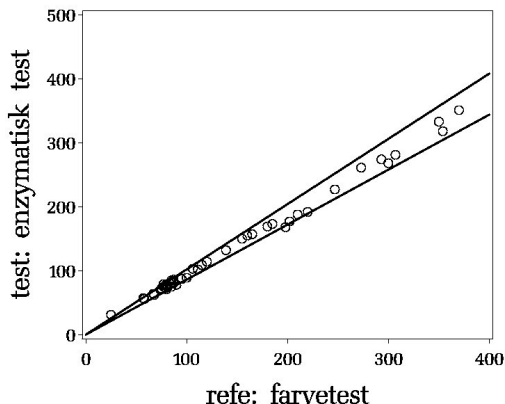
TEST ligger typisk mellem

14% under og **2% over** REFE.



Limits of agreement

tilbagetransformeret til oprindelig skala



Regninger direkte på ratio-skala

Med definitionen $\text{ratio} = \frac{\text{refe}}{\text{test}}$ får vi:

The MEANS Procedure

Variable	N	Mean	Std Dev	Std Error
ratio	45	1.0688607	0.0451184	0.0067259

svarende til **limits of agreement**:

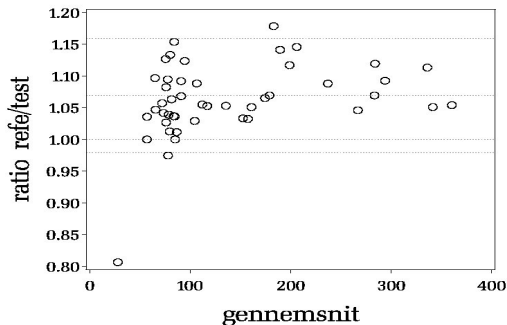
$$1.069 \pm 2 \times 0.045 = (0.979, 1.159)$$

altså refe fra 2% under til 16% over test,
på 2 decimaler identisk med resultatet for logaritmerne

Dette er ikke altid tilfældet!!



Limits of agreement på ratio-skala



som jo er næsten identisk med figuren s. 11 bortset fra skalaen på akserne.

Terminologi

for **kvantitativt outcome**, f.eks. vitamin D

Regressionsanalyse: Kovariaterne er også **kvantitative**

- ▶ Simpel (lineær) regression:
kun en enkelt kovariat
- ▶ Multipel (lineær) regression:
to eller flere kovariater

Variansanalyse: Kovariaterne er **kategoriske** (grupper)

- ▶ Ensidet variansanalyse: kun en enkelt kovariat
- ▶ Tosidet variansanalyse: to kovariater

Generel lineær model **GLM**:

Begge typer kovariater i samme model

- ▶ **Kovariansanalyse (ANCOVA)**:
Netop en kvantitativ og en kategorisk kovariat



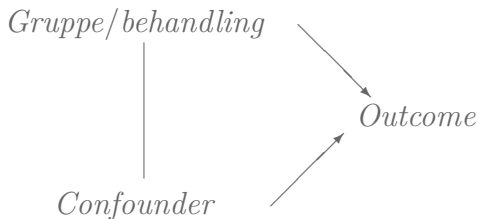
Sammenligning af to grupper

- som ikke er helt sammenlignelige, pga en **confounder**:

En variabel, som

- ▶ har en effekt på outcome
- ▶ er relateret til gruppen
(dvs. der er forskel på værdierne i de to grupper)

Herved **kan** opstå **“bias”**.



Eksempel: Vægt blandt mænd og kvinder

Mulig confounder: Højde

To forskellige sammenligninger:

- ▶ **T-test:**

Er der forskel på middel vægt blandt mænd og kvinder?

- ▶ **Kovariansanalyse:**

Er der forskel på middel vægt for mænd og kvinder,
med samme højde?

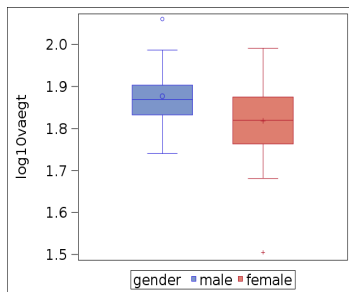
Kønseffekten *korrigeres* for forskel i højde:

To forskellige videnskabelige spørgsmål



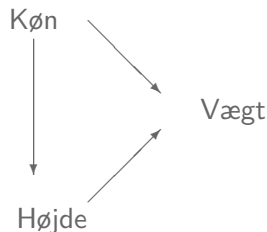
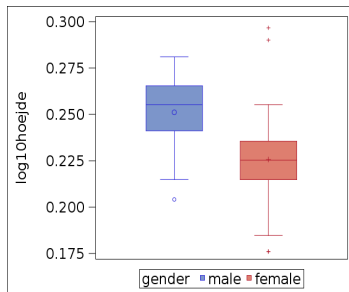
Vægtfordeling for mænd og kvinder

her for 100 tilfældigt udvalgte individer fra Sundby-materialet, på logaritmisk ($\log_{10}$ -skala):



Et T-test giver $P=0.0002$,
og en estimeret "*fordel*" til
mænd på 14.5% (6.8-22.8%),
se kode s. 83

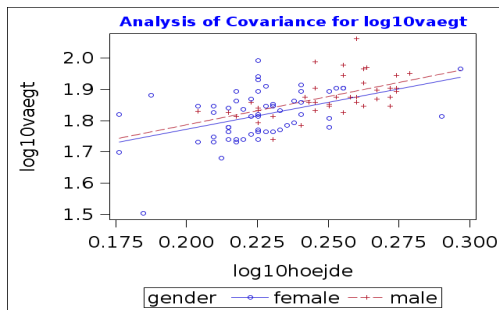
Kan højden forklare forskellen på mænd og kvinder?



- ▶ Der er i hvert fald forskel på mænds og kvinders højde
- ▶ Men er der tillige en ægte kønseffekt?

Relation mellem højde og vægt

Igen på $\log_{10}$ -skala, og med indlagte regressionslinier for hvert køn, se kode s. 84



Der er et par meget lave kvinder,
men ingen synderlige afvigelser fra linearitet...

Kovariansanalyse

Outcome: Vægt, logaritmeret ($\log_{10}\text{vaegt}$)

Kovariater: 2 stk:

- ▶ Køn (gender)
- ▶ Højde, logaritmeret ($\log_{10}\text{hoejde}$), med lineær effekt

```
proc glm data=sundby_udvalgt;  
  class gender;  
  model log10vaegt=log10hoejde gender  
    / solution clparm;  
run;
```



Output fra kovariansanalyse

Den vigtigste del:

Parameter		Estimate	Standard Error	t Value	Pr > t
Intercept		1.439299187 B	0.08236551	17.47	<.0001
log10hoejde		1.748543217	0.32637570	5.36	<.0001
gender	female	-0.017162755 B	0.01570514	-1.09	0.2773
gender	male	0.000000000 B	.	.	.

Parameter		95% Confidence Limits	
Intercept		1.275737609	1.602860765
log10hoejde		1.100425769	2.396660666
gender	female	-0.048350058	0.014024549
gender	male	.	.

Bemærk: Nu er kønseffekten helt forsvundet!

Højden kunne altså godt forklare den....måske



Fortolkning af kovariansanalysen

Vi fitter **to parallelle linier** (se s. 28) men både outcome og den kvantitative kovariat er logaritmeret.....

Fortolkning af kønseffekt:

På $\log_{10}$ -skala ligger mænds vægt 0.01716 højere end kvinders. Når vi tilbagetransformerer dette, får vi $10^{0.01716} = 1.040$, svarende til, at mænd vejer ca. 4% mere end kvinder **med samme højde**.

Tilsvarende tilbagetransformerer konfidensintervallerne:
 $(10^{-0.0140245}, 10^{0.04835}) = (0.968, 1.118)$, dvs. det kan tænkes, at mænds vægt er hele 11.8% højere end kvinders **med samme højde**, men det kan også tænkes, at den faktisk er 3.2% lavere (svarende til at gange med faktoren 0.968).



Fortolkning af kovariansanalysen, fortsat

Fortolkning af effekt af højde:

Højden på log-skala antages at have en lineær effekt på vægten på log-skala. Det betyder, at vi skal fortolke den tilbagetransformerede koefficient som *effekten af en 10-dobling af højden*, hvilket jo ikke er særligt interessant....

Når både outcome og kovariat er logaritmeret med samme logaritme, er det imidlertid ligegyldigt, hvad grundtallet for logaritmen er, da koefficienten altid vil blive den samme.

Vi kan derfor blot sige, at en 10% forøgelse af højden (svarende til faktoren 1.1) bevirker en faktor $1.1^{1.74854} = 1.181$ på vægten, altså en forøgelse af vægten med 18.1%.

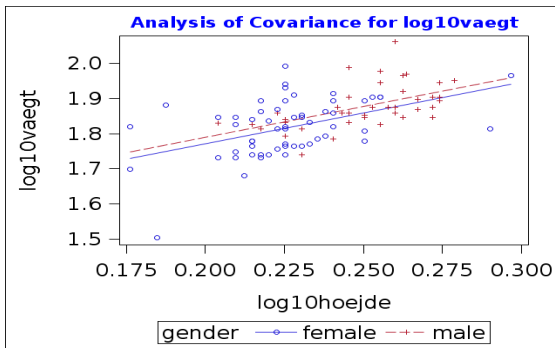
Tilsvarende tilbagetransformerer konfidensintervallerne:

$(1.1^{1.10043}, 1.1^{2.39666}) = (1.111, 1.257)$, dvs. mellem 11.1% og 25.7% mere i vægt.



Ancova-fit (parallelle linier)

Bemærk, at de parallelle linier er næsten sammenfaldende, svarende til den ringe kønseffekt.



Se kode s. 24

Estimation af kønseffekt

i 2 forskellige modeller, nemlig **med og uden højde** som kovariat

Outcome er $\log_{10}$ vægt:

Kovariater	Mænd vs. kvinder ratio (CI)	P-værdi
kun kønnet	1.14 (1.07, 1.23)	0.0002
køn og højde	1.04 (0.97, 1.12)	0.28

Den observerede forskel i ($\log_{10}$) vægt mellem mænd og kvinder **kan** altså evt. tilskrives højdeforskellen mellem kønnene.



Er der så en kønseffekt eller ej?

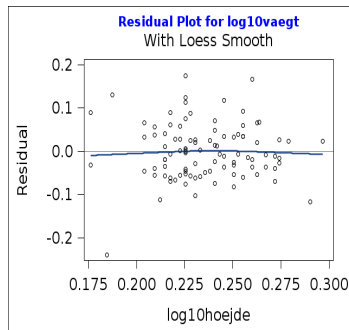
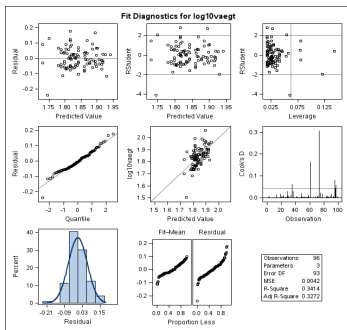
Ja for mænd og kvinder har jo ikke samme middel vægt
men forskellen i niveau kan forklares ud fra den
højdeforskel, der generelt er mellem mænd og kvinder
måske men vi kan ikke *påvise* forskel mellem en mand og en
kvinde med samme højde

Højde er en **mellemkommende** (intermediate) variabel
for effekten af køn på vægt



Husk modelkontrol:

Se kode s. 85



Additivitet – eller interaktion?

Kovariansanalyse: dækker sædvanligvis over situationen med to **parallelle** linier, altså med identiske hældninger, også kaldet **additivitet**

Her: Effekten af højde på vægt antages at være *den samme* for mænd og kvinder

Mere generel model tillader

interaktion=vekselvirkning=effektmodifikation, altså *forskellige* hældninger. Det betyder:

- ▶ Effekten af højde afhænger af kønnet
- ▶ Forskellen på kønnene afhænger af højden

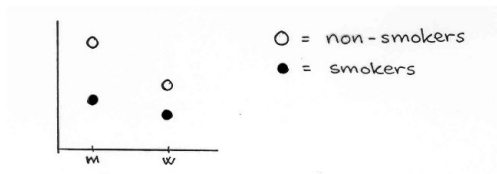
I tilfælde af interaktion kan man **ikke** udtale sig om **én** generel effekt af hverken højde eller køn.



Vekselvirkning (interaktion)

Tænkt eksempel:

- ▶ To inddelingskriterier: køn og rygestatus
- ▶ Outcome: FEV_1



- ▶ Effekten af rygning **afhænger af** køn
- ▶ Forskellen på kønnene **afhænger af** rygestatus

Mulige forklaringer

- ▶ **biologisk kønsforskel** på effekt af rygning
 - holder vist ikke i praksis,
men eksemplet er jo også blot 'tænkt'
- ▶ måske ryger kvinderne ikke helt så meget
 - antal pakkeår confounder for køn
- ▶ måske virker rygningen som en *relativ* (%-vis) nedsættelse af FEV_1
 - kunne undersøges ved en longitudinel undersøgelse



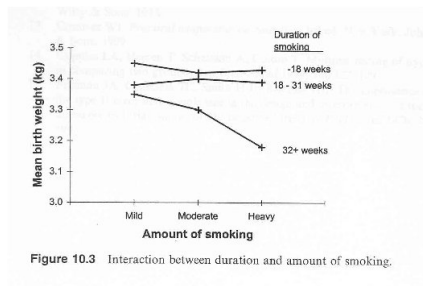
Eksempel: Rygnings effekt på fødselsvægt

Table 10.1 Average Birth Weight of Children Born to Women with Different Amount and Duration of Smoking^a

Duration of smoking in pregnancy	Amount of smoking			
	Mild	Moderate	Heavy	All
-18 weeks	3.45 (n = 15)	3.42 (n = 12)	3.43 (n = 7)	3.44 (n = 34)
18-31 weeks	3.38 (n = 8)	3.40 (n = 10)	3.39 (n = 6)	3.39 (n = 24)
32+ weeks	3.35 (n = 5)	3.30 (n = 3)	3.18 (n = 9)	3.25 (n = 17)
All	3.41 (n = 28)	3.40 (n = 25)	3.32 (n = 22)	3.38 (n = 75)

^a Entries are average birth weight in kg.





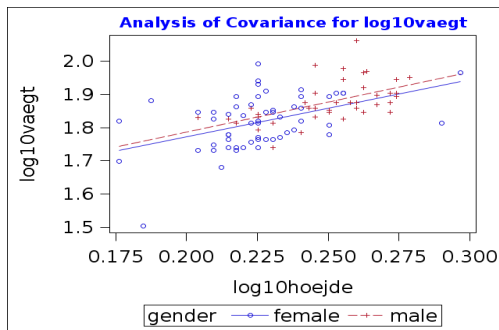
Interaktion/vekselvirkning
mellem mængden og varigheden af rygningen

- ▶ Der er effekt af mængden, men **kun** hvis man har røget længe.
- ▶ Der er effekt af varigheden, og denne effekt **øges** med mængden.

Effekten af mængden **afhænger af**....
og effekten af varigheden **afhænger af**....

Model med interaktion

Nu er linierne ikke mere *tvunget* til at være parallelle, men de er det næsten alligevel...



Kode s. 38



Model med interaktion i praksis

Testvenlig kode:

```
proc glm data=sundby_udvalgt;  
  class gender;  
  model log10vaegt=gender log10hoejde gender*log10hoejde  
    / solution clparm;  
run;
```

Mere estimat-venlig kode:

```
proc glm data=sundby_udvalgt;  
  class gender;  
  model log10vaegt=gender gender*log10hoejde log10hoejde  
    / noint solution clparm;  
run;
```



Output fra testvenlig model

Kode s. 38, foroven

Source	DF	Type I SS	Mean Square	F Value	Pr > F
gender	1	0.08277339	0.08277339	19.29	<.0001
log10hoejde	1	0.12186195	0.12186195	28.40	<.0001
log10hoejde*gender	1	0.00006689	0.00006689	0.02	0.9009

Source	DF	Type III SS	Mean Square	F Value	Pr > F
gender	1	0.00000208	0.00000208	0.00	0.9825
log10hoejde	1	0.11173743	0.11173743	26.04	<.0001
log10hoejde*gender	1	0.00006689	0.00006689	0.02	0.9009

Parameter		Estimate	Standard Error	t Value	Pr > t
Intercept		1.425145775 B	0.14038078	10.15	<.0001
gender	female	0.003697926 B	0.16782293	0.02	0.9825
gender	male	0.000000000 B	.	.	.
log10hoejde		1.805081993 B	0.55921280	3.23	0.0017
log10hoejde*gender	female	-0.086223040 B	0.69058249	-0.12	0.9009
log10hoejde*gender	male	0.000000000 B	.	.	.

Interaktionsparameteren angiver
forskellen på hældninger i de to grupper



Fortolkning af interaktionsmodel

Vi har fittet to *vilkårlige linier* (se s. 37) men stadig med logaritmeret outcome og kovariat.....

Fortolkning af kønseffekt:

Da linierne ikke er parallelle, er der ikke mere noget, der hedder kønseffekt~~en~~, da afstanden mellem linierne på s. 37 varierer en smule

Det angivne estimat for gender angiver forskellen på mænd og kvinder med $\log_{10} \text{hoejde} = 0$, dvs. en højde på 1 meter, altså *overhovedet ikke interessant*.

Fortolkning af interaktionsmodel, fortsat

Fortolkning af effekt af højde:

Nu er der *to forskellige effekter af højde*, svarende til de to køn. De fortolkes ligesom s. 27, idet koefficienten til $\log_{10}\text{hoejde}$ angiver effekten for referencegruppen for gender (altså mænd), medens koefficienten svarende til $\log_{10}\text{hoejde} * \text{gender female}$ angiver *forskellen* mellem effekten for kvinder og effekten for mænd.

Vi omregner disse estimater på s. 42, men se estimationsvenlig kode s. 38 (nederst) og output fra denne s. 43

Omregning til de to linier:

Linie for mænd:

$$\log_{10}(\text{vægt}) = 1.4251 + 1.8051 \times \log_{10}(\text{højde})$$

Linie for kvinder:

$$\begin{aligned} & \log_{10}(\text{vægt}) \\ = & 1.4251 + 0.0037 + (1.8051 - 0.0862) \times \log_{10}(\text{højde}) \\ = & 1.4288 + 1.7189 \times \log_{10}(\text{højde}) \end{aligned}$$

Men så får vi ikke nogen konfidensgrænser....
så man bør benytte programmet til udregningerne,
se kode nederst s. 38



Output fra estimationsvenlig kode

Kode s. 38 (nederst)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
gender	2	1.47808224	0.73904112	172.22	<.0001
log10hoejde*gender	1	0.00006689	0.00006689	0.02	0.9009
log10hoejde	1	0.11173743	0.11173743	26.04	<.0001

Parameter		Estimate	Standard Error	t Value	Pr > t
gender	female	1.428843701	0.09196615	15.54	<.0001
gender	male	1.425145775	0.14038078	10.15	<.0001
log10hoejde*gender	female	1.718858953 B	0.40519777	4.24	<.0001
log10hoejde*gender	male	1.805081993 B	0.55921280	3.23	0.0017
log10hoejde		0.000000000 B	.	.	.

Parameter		95% Confidence Limits	
gender	female	1.246190991	1.611496412
gender	male	1.146337436	1.703954114
log10hoejde*gender	female	0.914101208	2.523616698
log10hoejde*gender	male	0.694437112	2.915726874
log10hoejde		.	.



* Fortolkning af interaktionsmodel, fortsat

Vi har nu direkte estimater for de to linier (se s. 42-43) og skal tilbagetransformere til relationen

$$\text{vægt} = 10^{\alpha} \times \text{højde}^{\beta}$$

for hvert køn for sig. Vi finder:

► **For mænd:**

$$\text{vægt} = 10^{1.4251} \times \text{højde}^{1.8051} = 26.61 \quad \text{højde}^{1.8051}$$

► **For kvinder:**

$$\text{vægt} = 10^{1.4288} \times \text{højde}^{1.7189} = 26.84 \quad \text{højde}^{1.7189}$$

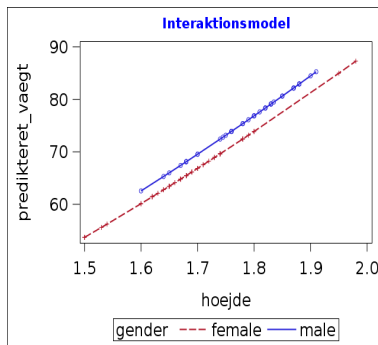
Disse *kurver* (**ikke linier**) er vist på s. 45.

Bemærk dog, at de ikke afviger ret meget fra linier....



Estimerede relationer i interaktionsmodellen

Da der er tale om linier på dobbeltlogaritmisk skala, er der ved tilbagetransformation tale om krumme kurver (dog ikke særligt krumme...)



Bemærk

Da der absolut ikke kunne spores nogen interaktion i dette eksempel, går vi tilbage til den additive model, *kovariansanalysen*, fra s. 24. I **dette** eksempel så vi

- Den observerede forskel i $(\log_{10})$ vægt mellem mænd og kvinder **kunne godt** tilskrives højdeforskellen mellem kønnene.

Der **kan** dog stadig være en kønsforskel **op til ca. 12% øget vægt hos mænd** i forhold til en kvinde på samme højde (se s. 29).

- Det *omvendte* kan også forekommer, altså at **effekter dukker op, når der kontrolleres for andre**

Sådan et eksempel skal vi se nu



Nyt eksempel: Hormoner hos P-pille brugere

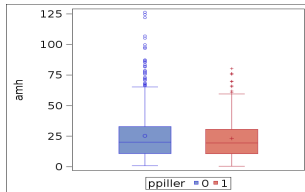
Anti Müllersk Hormon (**AMH**) er et hormon, som dannes i de små tidlige ægblærer (follikler) i æggestokkene.

Videnskabeligt spørgsmål:

Er AMH-niveauet nedsat hos P-pille brugere?

Og i givet fald hvor meget?

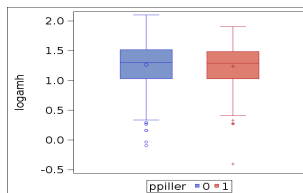
Der foreligger en undersøgelse af 732 kvinder, heraf 228, der tager P-piller



Valg af skala

- ▶ Boxplottet på forrige side viser klart en tendens til hale mod høje værdier
- ▶ Man får derfor en ide om, at logaritmer vil være påkrævet
- ▶ På den anden side er der mange observationer i hver gruppe, så det er muligvis ikke strengt nødvendigt....

Boxplot af logaritmetransformerede ($\log_{10}$) AMH-værdier



Heller ikke på denne skala er der perfekt symmetri

Valg af skala

er ikke altid oplagt....

Man kan lade det afhænge af, hvordan man helst vil præsentere sit resultat:

- ▶ som en forskel i enheden antal/ml, eller
- ▶ som en **relativ forskel**, i %

Hvis man vælger andre skalaer (kvadratrødder, kubikrødder etc.) kan parametrene i modellen ikke direkte fortolkes (forskellene kan ikke kvantificeres på en enkel måde).



Sammenligning af grupper - på logaritmisk skala

Uparret T-test giver:

The TTEST Procedure

Variable: logamh

ppiller	N	Mean	Std Dev	Std Err	Minimum	Maximum
0	504	1.2646	0.3655	0.0163	-0.0947	2.1014
1	228	1.2394	0.3626	0.0240	-0.4034	1.9043
Diff (1-2)		0.0252	0.3646	0.0291		

ppiller	Method	Mean	95% CL Mean	Std Dev
0		1.2646	1.2326 1.2966	0.3655
1		1.2394	1.1921 1.2867	0.3626
Diff (1-2)	Pooled	0.0252	-0.0319 0.0823	0.3646
Diff (1-2)	Satterthwaite	0.0252	-0.0318 0.0822	

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	730	0.87	0.3868
Satterthwaite	Unequal	441.56	0.87	0.3855

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	503	227	1.02	0.8987



Kommentarer til output

- ▶ Der er ikke tegn på forskellige varianser i de to grupper ($P=0.90$)
- ▶ Det er derfor ligegyldigt, hvilket T-test, vi vælger at se på (de giver i øvrigt også næsten præcis det samme)
- ▶ P-værdien er 0.39, altså **ingen evidens for forskel** i AMH blandt brugere og ikke-brugere af P-piller



Kommentarer til output,II

Måske er det mere naturligt at udregne det *den anden vej*, altså P-pille brugere i forhold til ikke P-pille brugere, og så skal vi skifte fortegn:

$$10^{-0.0252} = 0.94, \quad \text{CI} = (10^{-0.0823}, 10^{0.0319}) = (0.83, 1.08)$$

Her læser vi, at P-pille brugere i gennemsnit har et ca. 6% reduceret niveau af AMH, men at konfidensintervallet strækker sig fra en reduktion på ca. 17% til et forøget niveau på omkring 8%.

Det lyder som en **ret betydelig potentiel forskel** i mine ører....
men altså ikke signifikant

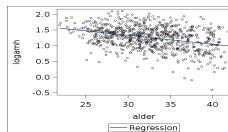
Kan der være en confounder? Alder?



Er alder en confounder?

- Har den en effekt på hormon-niveauet?

Det skulle man tro, ud fra vores viden om, hvad hormonet afspejler



- Er der forskellig alder i de to grupper?

Formentlig er P-pille brugerne yngre end ikke-P-pille brugerne

	N					
ppiller	Obs	Variable	N	Mean	Median	Std Dev
0	504	alder	504	33.7236111	33.7000000	4.0226765
1	228	alder	228	30.0162281	29.2000000	4.0995917

Alder er en confounder!

Kovariansanalyse - justering for aldersforskel

Nu sammenligner vi kvinder, der tager P-piller med kvinder, der *ikke* tager P-piller, men som har **samme alder**:

```
proc glm plots=(DiagnosticsPanel Residuals(smooth)) data=a1;
class ppiller rygning;
model logamh = alder ppiller / solution clparm;
output out=ny p=yhat r=residual cookd=cook;
run;
```

med output:

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	2.280455844 B	0.09508466	23.98	<.0001
alder	-0.034683747	0.00307944	-11.26	<.0001
ppiller 0	0.153783588 B	0.02919734	5.27	<.0001
ppiller 1	0.000000000 B	.	.	.

Parameter	95% Confidence Limits	
Intercept	2.093783409	2.467128278
alder	-0.040729371	-0.028638123
ppiller 0	0.096462683	0.211104494
ppiller 1	.	.



Kommentarer til kovariansanalyse

Kvinder, der tager P-piller estimeres til et 0.1538 lavere niveau end kvinder, der ikke tager P-piller, med $CI=(0.0965, 0.2111)$.

Når vi tilbagetransformerer til selve koncentrationsskalaen, giver det en estimeret ratio $10^{-0.1538} = 0.70$, med konfidensinterval $(10^{-0.2111}, 10^{-0.0965}) = (0.62, 0.80)$

Dette skal fortolkes som, at P-pille brugere ligger **30% lavere** end ikke-P-pille brugerne, med konfidensinterval fra 20% til 38% under.

Og forskellen er stærkt signifikant ($P < 0.0001$)



Kommentarer til de to analyser

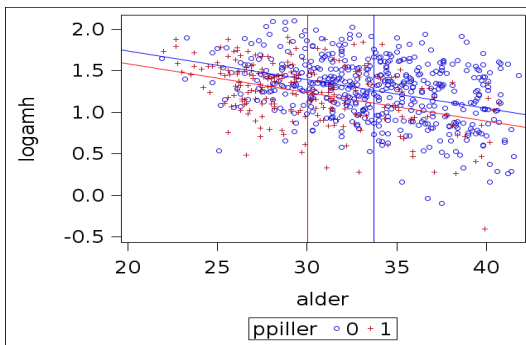
Vi sammenfatter i en tabel:

Model	Reduktion i AMH: P-pille brugere vs. ikke P-pille brugere	P-værdi
T-test (s. 50-52)	-6% (-17%, +8%)	0.39
Kovariansanalyse (s. 54-55)	-30% (-38%, -20%)	<0.0001



Kovariansanalyse modellen grafisk

med fortolkning næste side



Men kan der tænkes at være interaktion?

Fortolkning af figur s. 57

- ▶ Kovariansanalysen svarer til de to parallelle regressionslinier
- ▶ Den lodrette afstand imellem disse svarer til effekten af P-piller i kovariansanalysen s. 54-55
- ▶ De to lodrette linier er anbragt i aldersgennemsnittene:

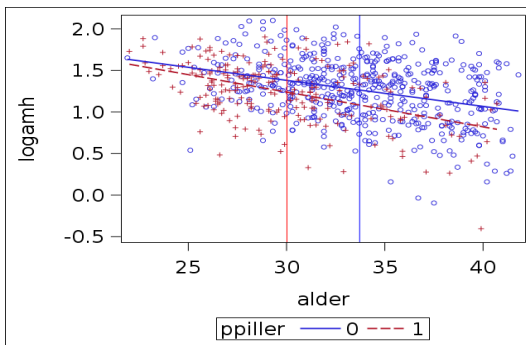
Rød P-pille brugere

Blå Ikke-P-pille brugere

Disse linier ses at skære regressionslinierne i *nogenlunde* samme højde (dvs. værdi af $\log_{\text{amh}}$), og ca. svarende til det insignifikante T-test s. 50-52

Figur med *frie* regressionslinier

svarende til *mulig interaktion*



Er de to linier ikke-parallelle?

Fortolkning af figur s. 59

- ▶ Interaktionsmodellen svarer til de to *ikke*-parallelle regressionslinier
- ▶ Den lodrette afstand imellem disse varierer nu med alderen, så man ikke mere kan tale om en *generel* effekt af P-piller
- ▶ De to lodrette linier skærer nu regressionslinierne i *præcis* de gennemsnitlige værdier af $\log_{10} h$, svarende til det insignifikante T-test s. 50-52

Er effekten af alder forskellig i de to grupper?

Regression for hver gruppe for sig (kode s. 90)

ppiller=0

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	2.320725627	0.12927526	17.95	<.0001
alder	-0.031317743	0.00380644	-8.23	<.0001

Parameter	95% Confidence Limits	
Intercept	2.066738419	2.574712835
alder	-0.038796267	-0.023839219

ppiller=1

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	2.496012716	0.15697593	15.90	<.0001
alder	-0.041865092	0.00518181	-8.08	<.0001

Parameter	95% Confidence Limits	
Intercept	2.186689089	2.805336343
alder	-0.052075923	-0.031654260

Konfidensintervallerne for alderseffekterne er noget overlappende, så vi kan ikke umiddelbart udtale os om, hvorvidt de er signifikant forskellige.



Interaktion mellem P-piller og alder?

De to analyser s. 61 viser, at effekten af 5 år (5 gange hældningen) kan opsummeres som i nedenstående tabel, der også tilbagetransformerer effekten til original skala:

Gruppe	Effekt af 5 år på $\log_{10}$ -skala	Tilbagetransformeret effekt
Ikke P-pille brugere	-0.1566 (-0.1940, -0.1192)	0.70 (0.64, 0.76)
P-pille brugere	-0.2093 (-0.2604, -0.1583)	0.62 (0.55, 0.69)

Fortolkningen er, at for P-pille brugere falder AMH-niveauet med 38% på 5 år (faktoren 0.62), medens der for ikke P-pille brugere kun er tale om et fald på 30% (faktoren 0.7).

Interaktion mellem P-piller og alder, II

Er der evidens for reelle forskelle i alderseffekten i de to grupper?

For at undersøge dette, skal vi først gøre os klart, at spørgsmålet går på, om de to hældninger ovenfor er signifikant forskellige, altså om alderseffekten afhænger af, om man er P-pille bruger eller ej.

```
proc glm plots=(DiagnosticsPanel Residuals(smooth)) data=a1;
class ppiller rygning;
model logamh = alder ppiller ppiller*alder / solution clparm;
output out=ny p=yhat r=residual cookd=cook;
run;
```

Parameter		Estimate	Standard Error	t Value	Pr > t
Intercept		2.496012716 B	0.16495869	15.13	<.0001
alder		-0.041865092 B	0.00544532	-7.69	<.0001
ppiller	0	-0.175287089 B	0.20794681	-0.84	0.3995
ppiller	1	0.000000000 B	.	.	.
alder*ppiller	0	0.010547348 B	0.00659921	1.60	0.1104
alder*ppiller	1	0.000000000 B	.	.	.



Kommentarer til output vedr. interaktion

- ▶ Testet for ens hældninger aflæses på linien `alder*ppiller`, og P-værdien ses at være 0.11, altså ingen signifikans.

I hvert fald ikke, hvis vi holder os til et 5% signifikansniveau.

- ▶ Vi kan således ikke i dette materiale påvise en forskel i alderseffekt for de to grupper, omend der synes at være tendens til en større effekt i P-pille gruppen.

Der *kunne* således være en forskel i hældninger på op imod $0.0105 + 2 \times 0.0066 = 0.0237$ svarende til, at P-pille brugere på 5 år havde et ekstra tab på 24% (fordi $10^{-5 \times 0.0237} = 0.76$)



Kvantificering i tilfælde af interaktion

Her finder vi ingen signifikant interaktion mellem alder og indtagelse af P-piller, men dette *udelukker* jo ikke, at der kan være interaktion.

Og hvordan kvantificerer man så effekten af P-piller i tilfælde af interaktion?

- ▶ Man kan kvantificere effekten for udvalgte aldre, f.eks. 25, 30, 35 og 40, ved hjælp af estimate-sætninger, se s. 66
- ▶ eller ved at ændre Y-aksens placering ved at flytte nulpunktet, så effekten kvantificeres som forskellen i intercepter

Selve niveauerne kan kvantificeres for en lang række aldre ved at inkludere disse i datasættet med *missing* i AMH-outcome, se 92



Kvantificering med *estimate*-sætninger

```
proc glm data=a1;
class ppiller rygning;
model logamh = ppiller ppiller*alder / solution clparm;
estimate 'effekt af P-piller, 25 aar'
  ppiller -1 1 alder*ppiller -25 25;
estimate 'effekt af P-piller, 30 aar'
  ppiller -1 1 alder*ppiller -30 30;
estimate 'effekt af P-piller, 35 aar'
  ppiller -1 1 alder*ppiller -35 35;
estimate 'effekt af P-piller, 40 aar'
  ppiller -1 1 alder*ppiller -40 40;
run;
```

med det ekstra output

Parameter	Estimate	Standard Error	t Value	Pr > t
effekt af P-piller, 25 aar	-0.08839662	0.05024320	-1.76	0.0789
effekt af P-piller, 30 aar	-0.14113336	0.03022113	-4.67	<.0001
effekt af P-piller, 35 aar	-0.19387010	0.03846730	-5.04	<.0001
effekt af P-piller, 40 aar	-0.24660684	0.06498945	-3.79	0.0002

som så skal tilbagetransformeres.....



Metoder til at undgå bias (her pga alder)

Matchning. Dvs. udvælge individer, således at de er nogenlunde ens med hensyn til de vigtige *forstyrrende* kovariater.

Husk at tage matchningsvariablen med som kovariat

Randomisering. Dvs. trække lod om behandling (gruppe)

NB: Dette kan naturligvis *kun* lade sig gøre, hvis grupperne er noget, man selv bestemmer over.

og det sikrer ikke helt identiske fordelinger i den enkelte undersøgelse

Korrektion Dvs. at medtage den (muligvis skævt fordelte) variabel som kovariat, **også somme tider i randomiserede undersøgelser**



Matching / randomisering / korrektion

► Matching:

- god metode, når den ene gruppe er vanskelig at få tag i
- kan være ret besværligt og kan reducere antallet af observationer
- skævvrider udtrykket for populationen som helhed

► Randomisering:

- vældig god metode, når den er mulig og sample size er stor
- vi kan ikke randomisere til P-piller ja/nej

► Korrektion:

- er altid mulig, men problematisk hvis der er for stor forskel på grupperne
- kan reducere residualvariationen (væsentligt)
- sample size bestemmer hvor meget, det er muligt at korrigere for....



Metoder til at øge styrken

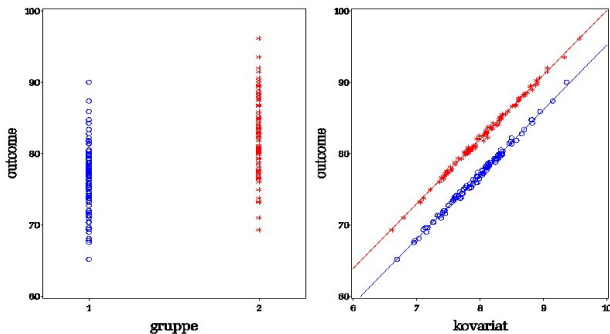
- ▶ inkludere flere observationer/personer
- ▶ benytte fornuftige inklusions- eller eksklusionskriterier
- ▶ inddrage vigtige forklarende variable (kovariater)
også selv om de ikke er confoundere,
se eksempel på en sådan næste side
- ▶ udelade irrelevante kovariater

Men pas på med at gå for meget på **fisketur!!**

og husk at **fortolkningen afhænger af modellen**



Simuleret eksempel



Uden x i modellen: Ingen særlig forskel på grupperne...?

Med x i modellen: Tydelig forskel på grupperne
(her den lodrette afstand mellem linierne)

Før-og-efter studie

Et typisk eksempel:

Vickers, A.J. & Altman, D.G.: Analysing controlled clinical trials with baseline and follow-up measurements.

British Medical Journal 2001; **323**: 1123-24.:

52 patienter med skuldersmerter **randomiseres** til enten

- ▶ Akupunktur (n=25)
- ▶ Placebo (n=27)

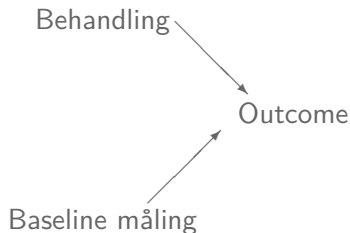
Smerte vurderes på en 100-trins skala
både før og efter behandling

Høje scorer er gode



Baseline og Follow-up

I princippet:



Der er fokus på sammenligning af behandlingerne:

Er placebo ligeså godt som akupunktur?

Det lægger op til et simpelt T-test, **men:**

Hvad nu, hvis de to grupper ikke er *helt* identiske fra start, selv om der er randomiseret?

Gentagne målinger over tid

– egentlig først emnet den sidste uge af kurset, men her er det simple, fordi vi kun har 2 målinger pr. individ, så:

Vi kan se på ændringerne, men er dette fornuftigt?

Ikke altid, pga regression to the mean eller “reversion” towards the mean, som kan oversættes til

en tendens til at falde ind mod midten af en fordeling

Talrige eksempler fra dagligdagen:

- ▶ Efterfølgeren til en rigtig god film bliver sjældent helt så god
- ▶ Lungekræftdødeligheden i Fredericia er aftaget
- ▶ Placebo-effekten

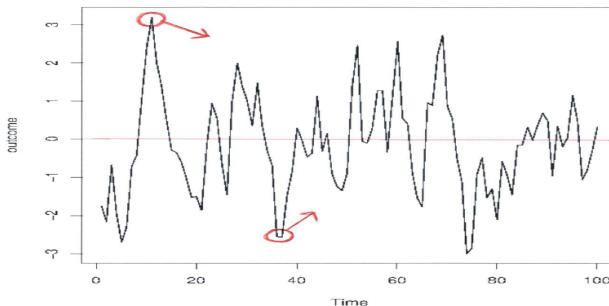
Det handler om **seleksion**



Hypotetisk situation

Stabilt outcome:

Hvordan udvikler folk sig, hvis de udvælges, når de udviser en ekstrem værdi?

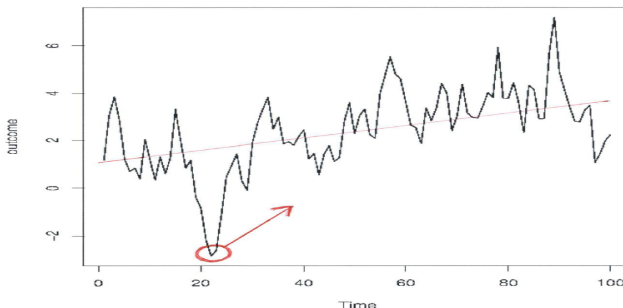


- ▶ Placebo-behandling af folk med forhøjet blodtryk virker....
- ▶ Folk med for lav hæmoglobin har godt af at tage ekstra

Hypotetisk BMD el.lign.

Outcome med positiv trend:

Hvordan udvikler folk sig, hvis de tilfældigvis udvælges, når de har et lavt niveau?



Den gruppe, der tilfældigvis ligger lavt fra starten, vil have tendens til den største stigning.

Hvor vigtigt er det?

Selektionseffektion er størst, når korrelationen mellem gentagne målinger (ρ) er *lille*,
men hvad betyder det?

Det betyder, at vi har at gøre med
et outcome, der kan *variere hurtigt*, f.eks.

- ▶ blodtryk
- ▶ hæmoglobin?
- ▶ smerter

men ikke

- ▶ højde
- ▶ knogletæthed



Resultater vedr. skuldersmerter

Sammenligninger af de to grupper:

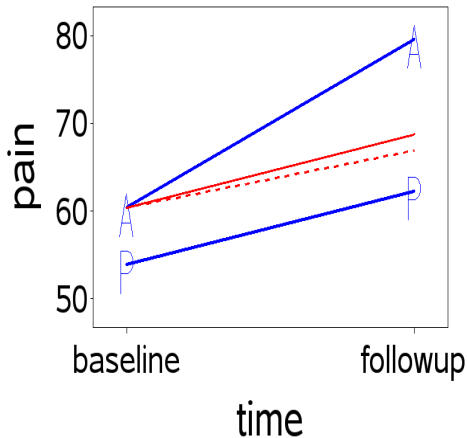
	Gennemsnitlig smertescore placebo (n=27)	akupunktur (n=25)	Behandlinseffekt differens (95% CI)	P-value
Baseline	53.9 (14.0)	60.4 (12.3)	6.5	0.09
Analyse:				
Follow-up	62.3 (17.9)	79.6 (17.1)	17.3 (7.5; 27.1)	0.0008
Ændringer*	8.4 (14.6)	19.2 (16.1)	10.8 (2.3; 19.4)	0.014
Ancova			12.7 (4.1; 21.3)	0.005

* resultater publiceret i Kleinhenz et.al. *Pain* 1999; 83:235-41.

De to baseline værdier er ikke signifikant forskellige,
men de er heller ikke helt ens....



Udvikling i smerter, faktisk og forventeligt



Udlægning af resultater, I

Baseline

- ▶ Akupunktur gruppen ligger lidt højere end placebo

Follow-up

- ▶ Vi vil forvente, at akupunktur gruppen stadig ligger det samme stykke højere end placebo gruppen efter behandlingen, selv hvis behandlingen ikke virker (fuldt optrukket rød linie).
- ▶ Det er derfor urimeligt blot at afgøre behandlingens effekt ved at sammenligne follow-up værdier (forskellen bliver for stor i dette tilfælde)



Udlægning af resultater, II

Differenser/ændringer

- ▶ Lav baseline værdi giver forventning om stor positiv ændring (**regression to the mean**)
- ▶ Derfor forventes placebogruppen at stige mest, og akupunkturgruppen lidt mindre (**den stiplede røde linie**)
- ▶ og en direkte sammenligning af ændringer er derfor ikke rimelig (forskellen bliver for lille i dette tilfælde)

Ancova

- ▶ vurderer forskellen under hensyntagen til, at baseline forklarer *noget*, men *ikke al* forskellen ved follow-up:
Vi sammenligner personer, der har fået forskellig behandling, men som **starter samme sted**.



APPENDIX

Programbidder svarende til diverse slides:

- ▶ Bland-Altman plot, s. 82
- ▶ T-test, s. 83
- ▶ Scatterplot. s. 84
- ▶ Kovariansanalyse, s. 85-??, 89
- ▶ Model med interaktion, s. 87-88, 91
- ▶ Opdelte analyser, s. 90



Bland-Altman plot

Slide 8

```
data a1;
input refe test;
datalines;
155 150
160 155
180 169
80 79
/* flere datalinier her */
;

data a2;
set a1;

snit=(refe+test)/2;
dif=refe-test;
run;

proc gplot data=a2;
  plot dif*snit
    / vref=0 lv=2 vref=9.89 vref=-9.51 vref=29.29 lv=33
      haxis=axis1 vaxis=axis2 frame;
axis1 value=(H=2) minor=NONE
      label=(H=3 'gennemsnit');
axis2 value=(H=2) minor=NONE
      label=(A=90 R=0 H=3 'differens refe-test');
symbol1 v=circle i=none c=BLACK h=2 l=1 w=2;
run;
```



T-test på logaritmer

Slide 21

```
ods graphics on;  
proc ttest data=sundby_udvalgt;  
  class gender;  
  var log10vaegt;  
run;  
ods graphics off;
```



Scatter plot, med regressionslinier

Slide 23

```
proc sgplot data=sundby_udvalgt;  
    reg x=log10vaegt y=log10hoejde / group=gender;  
run;
```



Kovariansanalyse - ANCOVA

Slide 24, 28 og 31

Koden nedenfor giver automatisk noget modelkontrol (se dog også s. 86), samt plottet med de to parallelle linier:

```
ods graphics on;  
proc glm plots=all data=sundby_udvalgt;  
  class gender;  
  model log10vaegt=log10hoejde gender  
    / solution clparm;  
run;  
ods graphics off;
```



Modelkontrol i ANCOVA

Slide 31

For at få begge figurer s. 31 skrives det nedenfor anførte i `plots=`:

```
ods graphics on;  
proc glm plots=(diagnosticspanel residuals(smooth)) data=sundby_udvalgt;  
  class gender;  
  model log10vaegt=log10hoejde gender  
        / solution clparm;  
output out=ny p=predicted r=residual cookd=cook;  
run;  
ods graphics off;
```

Hvis man ønsker fuld kontrol over modelkontrollen, kan man danne et nyt datasæt med udvalgte ekstra størrelser, som vist i output-sætningen ovenfor.



Model med to regressionslinier

altså **inkluderende interaktion**

Slide 37

```
ods graphics on;  
proc glm plots=all data=sundby_udvalgt;  
    class gender;  
    model log10vaegt=gender gender*log10hoejde log10hoejde  
        / noint solution clparm;  
run;  
ods graphics off;
```



Omregning til to linier

Slide 42-43

1. Den lette metode:

```
proc glm data=sundby_udvalgt;  
  class gender;  
  model log10vaegt=gender gender*log10hoejde log10hoejde  
    / noint solution clparm;  
run;
```

2. Den fleksible metode:

```
proc glm data=sundby_udvalgt;  
  class gender;  
  model log10vaegt=gender log10hoejde gender*log10hoejde  
    / solution clparm;  
  estimate 'intercept, male' intercept 1 gender 0 1;  
  estimate 'intercept, female' intercept 1 gender 1 0;  
  estimate 'slope, male' log10hoejde 1 gender*log10hoejde 0 1;  
  estimate 'slope, female' log10hoejde 1 gender*log10hoejde 1 0;  
run;
```



Kovariansanalyse

Slide 54

```
ods graphics on;  
proc glm plots=(DiagnosticsPanel Residuals(smooth)) data=a1;  
class ppiller rygning;  
model logamh = alder ppiller / solution clparm;  
output out=ny p=yhat r=residual cookd=cook;  
run;  
ods graphics off;
```



Opdelte analyser

Slide 61

```
proc sort data=a1; by ppiller;  
run;  
  
ods graphics on;  
proc glm plots=all data=a1; by ppiller;  
model logamh=alder / clparm;  
estimate "fald paa 5 aar" alder 5;  
estimate "niveau ved alder 30" intercept 1 alder 30;  
run;  
ods graphics off;
```



Test for interaktion

Slide 63

```
ods graphics on;  
proc glm plots=(DiagnosticsPanel Residuals(smooth)) data=a1;  
class ppiller rygning;  
model logamh = alder ppiller ppiller*alder / solution clparm;  
output out=ny p=yhat r=residual cookd=cook;  
run;  
ods graphics off;
```



Prediktion i tilfælde af interaktion

Slide 65

Hvis vi vil prediktere niveauet for begge grupper for hver hele alder fra 20 til 40 (21 aldre), tilføjer vi 2×21 observationer:

```
alder ppiller amh
20      0      .
21      0      .
22      0      .
osv
osv
40      0      .
20      1      .
osv
osv
39      1      .
40      1      .
```

Outcome $\log_{\text{amh}}$ vil da blive predikteret for hver af disse personer, og man kan derefter fratrække for samme alder, men forskellig P-pille status og lave en figur.

